15th International Roundtable on Business Survey Frames
Washington, D.C. – October 22 – 26, 2001

Session No 2
Paper No 17
[*Richard CLAYTON*], [Bureau of Labor Statistics],[*United States*]

Progress Report

1. Introduction

The driving force for the Bureau of Labor Statistics' (BLS) business register has been the conversion to the new North American Industry Classification System (NAICS). The NAICS system is replacing the Standard Industrial Classification (SIC) system that has been in use since the 1930s.

A second major initiative is the completion of a longitudinal database providing linked quarterly business records covering the period 1990 to the present. This rich source of new data will be used to explore such economic issues as business birth and death measures, job creation and destruction, business survival studies, and others.

Thirdly, the tragic events in New York of September 11 lead to a rapid response analysis of the employment in the affected areas. The resulting analysis, conducted using postal zip codes emphasised the need for the business file to be geocoded to allow more precise geographic boundaries to be defined matching those of the area affected by an economic event. A separate paper on this topic is presented in a later session.

2. NAICS

The North American Industry Classification System (NAICS) is replacing the traditional Standard Industrial Classification System (SIC). The SIC system is based on grouping businesses by their what they produce, whereas NAICS groups businesses on a production or supply function. That is, establishments using similar raw material inputs, capital equipment and labor are classified in the same industry. Also, a number of new industries were identified to reflect the current economy, with an emphasis on the tremendous growth in service industries and to capture new advanced technology industries.

The conversion to the NAICS structure required the contact for each of the 8 million businesses in the United States with workers covered by Unemployment Insurance (UI). Over 3 years, different portions of the universe were mailed forms showing their SIC code and its description. A number of NAICS detailed descriptions were presented allowing the respondent to select the appropriate detailed industry, or to use provided space to write down a description of their processes and output, allowing subsequent coding by trained experts.

The three-year effort yielded overall coding rates of 97.3 percent of units and 99.7 percent of employment. This response provides a strong basis for the establishment-based surveys in BLS to revise their samples to represent the NAICS-based universe. Ongoing quality studies will review consistency and accuracy of the coding across the states and industries.

The original NAICs structure was published in 1998, called NAICS 1997. Continuing efforts to complete the work in selected sectors, notably Wholesale Trade, Retail Trade and Information, yielded the NAICS 2002 version. BLS will only publish on the NAICS 2002 version, skipping NAICS 1997, relieving users from such rapid and repeated breaks in series.

Published data using the NAICS structure will begin with the 2001 reference period. Ratio tables will be published in early 2002 to allow users to review the shifts in the employment and wages caused by the conversion to NAICS. About 68% on industries are direct conversions from SIC to NAICS.

Historical data is needed by the many users and analysts who rely on the long historical series for modelling and projections. The ES-202 program is currently planning to develop national level data on NAICS going back to 1990. In addition, state and county data is being reconstructed and will be reviewed to see where it can be published without risking disclosures of confidential data. This is particularly difficult because it may be possible to cross match published detailed SIC data versus detailed NAICS-based data to isolate a single business and thus reveal its information. Techniques for reducing this risk are being developed.

The historical data will also be used to provide a structure for the monthly payroll employment survey, the Current Employment Statistics program to link its historical data and recalculate their series much further back, in some cases to 1939.

3. Longitudinal Establishment Database

The demand for longitudinal data files is growing as a tool for deepening our understanding of economic events and phenomena. The ES-202 program completed the design and loading of a file containing linked records covering the period from 1990 to present. The resulting file, called the Longitudinal Database (LDB), is being used for a series of studies. Topics include gross job flows (job creation and destruction, and job reallocation) at varying geographic and industry levels, firm survival studies, effects of changes to the minimum wage, industrial location and clustering, the quantification of the non-profit industry, and economic impact studies such as military base closings or other local events. Researchers can apply to use this file at BLS.

New data for each quarter are loaded when review, editing, and linking is complete. There are two pieces of information in the ES-202 microdata that allow for matching establishments across quarters. First, is a unique identifier that each state assigns. Second, predecessor and successor numbers are assigned that allow tracking units that change sell or buy other units. Almost all establishments identified as continuous from quarter to quarter are matched by the unique identifier. Although, the small percentage linked by the predecessor/successor have a significant effect on the number of births and deaths.

Including the government sector, the LDB has over 8 million records for the current quarter and over 14 million establishments back to 1990. For the current year, there were approximately 1.45 million births and 1.3 million deaths, which is an increase of approximately 150,000 establishments per year. One of the most important characteristics of the LDB is the ability to quantify job flows. It provides the capability to track changes in employment and wages not only at the macro level but also at the micro level of the establishments. These job creation and destruction statistics have the potential to be an important new source for expanding our understanding of labor market demands.

The LDB is used to generate high quality, high frequency, timely and historical consistent information regarding job creation and destruction, and the life cycle of establishments. Eventually, quarterly and annual tables of job flows and establishment flows will be created and published.

4. Geocoding

Geocoding is the assignment of precise latitude and longitude coordinates to objects on the surface of the earth. The demand for new data to better understand economic phenomena is being met by the study and development of geocodes for the ES-202 program that houses the BLS business register. A more complete description of the research findings is provided later in this conference.